

EUROPEAN PATENT OFFICE

Patent Abstracts of Japan

PUBLICATION NUMBER : 2001236089
PUBLICATION DATE : 31-08-01

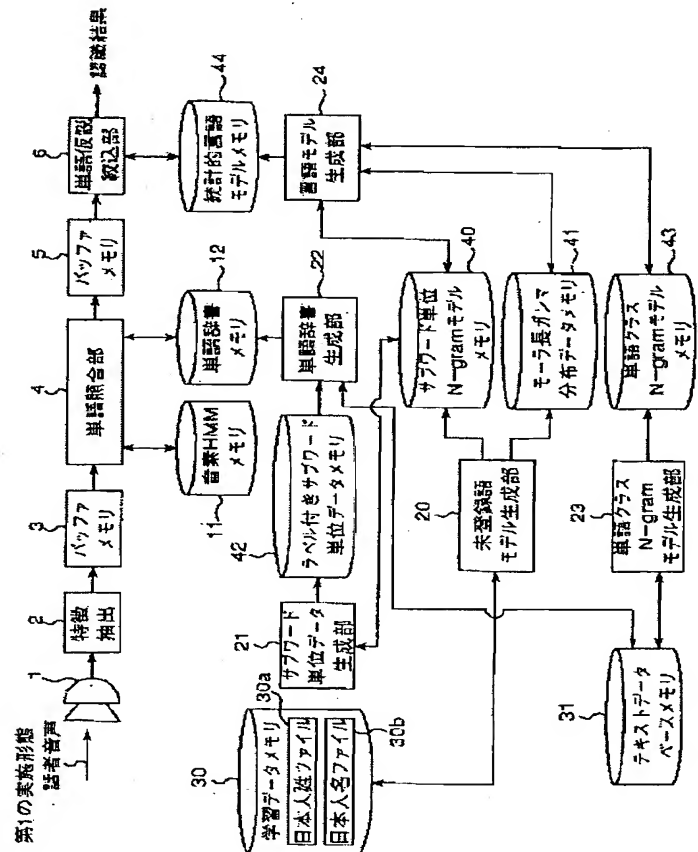
APPLICATION DATE : 13-12-00
APPLICATION NUMBER : 2000378702

APPLICANT : ATR INTERPRETING
TELECOMMUNICATIONS RES LAB;

INVENTOR : KOSAKA YOSHINORI;

INT.CL. : G10L 15/18 G06F 17/28

TITLE : STATISTICAL LANGUAGE MODEL
GENERATING DEVICE, SPEECH
RECOGNITION DEVICE,
INFORMATION RETRIEVAL
PROCESSOR AND KANA/KANJI
CONVERTER



ABSTRACT : PROBLEM TO BE SOLVED: To generate a statistical language model capable of enhancing the precision of speech recognition with respect to an unregistered word in a word dictionary and identifying the domain and class of the unregistered word.

SOLUTION: An unregistered word model generating section 20 assures that the ratio of the number of words to a mora length in learning data is practically defined as a gamma distribution and estimates and computes the parameters of the gamma distribution of mora lengths while depending on classes, computes the appearance probability of first N-gram which has the class that is a low-order class of a proper noun or a common noun of an adopted word in a subword unit that is mora or a mora link and generates a subword unit N-gram model which is made by modeling word series including unregistered words. A language model generating section 24 generates a statistical language model including unregistered words based on the subword unit based on the word class N-gram model and the subword unit N-gram model and the parameters of a gamma distribution of a mora length.

COPYRIGHT: (C)2001,JPO

(11)特許出願公開番号

特開2001-236089

(P2001-236089A)

(43)公開日 平成13年8月31日(2001.8.31)

| (51)Int.Cl. ⁷ | 識別記号 | F I | ページ* (参考) |
|--------------------------|------|---------------|-----------|
| G 1 0 L 15/18 | | G 0 6 F 17/28 | V |
| G 0 6 F 17/28 | | G 1 0 L 3/00 | 5 3 7 D |

審査請求 有 請求項の数10 O L (全 22 頁)

| | | | |
|--------------|-----------------------------|----------|--|
| (21) 出願番号 | 特願2000-378702(P2000-378702) | (71) 出願人 | 593118597 株式会社エイ・ティ・アール音声翻訳通信 研究所 京都府相楽郡精華町光台二丁目2番地2 |
| (22) 出願日 | 平成12年12月13日(2000. 12. 13) | (72) 発明者 | 谷垣 宏一 京都府相楽郡精華町大字乾谷小字三平谷5 番地 株式会社エイ・ティ・アール音声翻 訳通信研究所内 |
| (31) 優先権主張番号 | 特願平11-358947 | (74) 代理人 | 100062144 弁理士 青山 葆 (外2名) |
| (32) 優先日 | 平成11年12月17日(1999. 12. 17) | | |
| (33) 優先権主張国 | 日本(JP) | | |

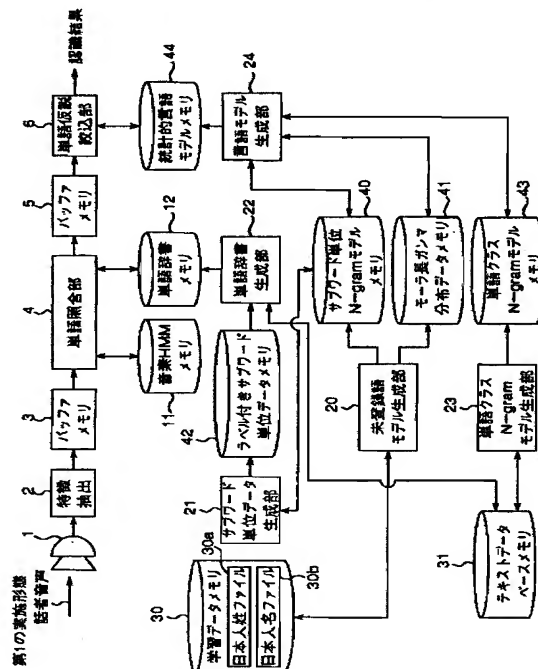
[最終頁に続く](#)

(54) 【発明の名称】 統計的言語モデル生成装置、音声認識装置、情報検索処理装置及びかな漢字変換装置

(57) 【要約】

【課題】 単語辞書において未登録の未登録語に関する音声認識の精度を高くし、未登録語の区間やクラスを同定する統計的言語モデルを生成する。

【解決手段】 未登録語モデル生成部20は学習データにおけるモーラ長に対する単語数の割合が実質的にガンマ分布に従うと仮定したときのモーラ長のガンマ分布のパラメータをクラスに依存して推定して計算し、モーラ又はモーラ連鎖であるサブワード単位で、固有名詞又は外来語の普通名詞の下位クラスであるクラスを有する第1のN-gramの出現確率を計算して未登録語を含む単語系列をモデル化したサブワード単位N-gramモデルを生成する。言語モデル生成部24は、単語クラスN-gramモデルとサブワード単位N-gramモデルとモーラ長のガンマ分布のパラメータとに基づいてサブワード単位に基づいた未登録語を含む統計的言語モデルを生成する。



【特許請求の範囲】

【請求項1】 クラス別に分類された単語リストを含む学習データメモリと、

その学習データに基づいて単語の読みに対応するサブワード系列の生起確率をクラスに依存して抽出する手段と、

その抽出された単語の読みに対応するサブワード系列の生起確率をクラスに依存して評価する手段と、

単語とクラスの並びの生起確率を評価する手段と、

上記クラスに依存して評価されたサブワード系列の生起確率と上記評価された単語とクラスの並びの生起確率とによって未登録語を含む単語系列の生起確率をサブワードと単語の系列の生起確率として評価する手段と、

を備えたことを特徴とする統計的言語モデル生成装置。

【請求項2】 上記サブワード系列の生起確率は、長さに関する確率と並びに関する確率とを含むことを特徴とする請求項1記載の統計的言語モデル生成装置。

【請求項3】 上記サブワードとしては、クラスに依存して抽出する単位を用いることを特徴とする請求項1又は2記載の統計的言語モデル生成装置。

【請求項4】 入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項1、2又は3記載の統計的言語モデル生成装置によって生成された統計的言語モデルを用いて音声認識することを特徴とする音声認識装置。

【請求項5】 固有名詞又は外来語の普通名詞の単語リストを含む学習データを格納する学習データ記憶手段と、

上記学習データ記憶手段に格納された学習データに基づいて、上記学習データにおけるモーラ長に対する単語数の割合が実質的にガンマ分布に従うと仮定したときのモーラ長のガンマ分布のパラメータをクラスに依存して推定して計算するとともに、モーラ又はモーラ連鎖であるサブワード単位で、上記固有名詞又は外来語の普通名詞の下位クラスであるクラスを有する第1のN-gramの出現確率を計算することにより未登録語をモデル化したサブワード単位N-gramモデルを生成する第1の生成手段と、

所定のテキストデータベースに基づいて生成された単語クラスN-gramモデルと、上記第1の生成手段によって生成されたサブワード単位N-gramモデルと、上記第1の生成手段によって計算されたモーラ長のガンマ分布のパラメータとに基づいて、上記単語クラスと、上記固有名詞又は外来語の普通名詞の下位クラスであるクラスとに依存した第2のN-gramの出現確率を計算することによりサブワード単位に基づいた未登録語を含む統計的言語モデルを生成する第2の生成手段とを備えたことを特徴とする統計的言語モデル生成装置。

【請求項6】 上記第1の生成手段によって生成されたサブワード単位N-gramモデルに基づいて、上記サブワード単位を抽出し、抽出したラベルを上記サブワード単位に付与することにより、サブワード単位当たり複数のラベル付きサブワード単位のデータを生成する第3の生成手段と、

上記テキストデータベースから抽出された単語と、上記第3の生成手段によって生成された複数のラベル付きサブワード単位のデータとに対して音素並びを付与することにより単語辞書を生成する第4の生成手段とをさらに備えたことを特徴とする請求項5記載の統計的言語モデル生成装置。

【請求項7】 入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項5又は6記載の統計的言語モデル生成装置によって生成された統計的言語モデルと、請求項6記載の第4の生成手段によって生成された単語辞書とを用いて音声認識することを特徴とする音声認識装置。

【請求項8】 上記単語リストに対応する普通名詞の単語データとそれに対応する情報とを含むデータベースを記憶するデータベース記憶手段と、

請求項4又は7記載の音声認識装置から出力される音声認識結果の文字列をキーとして用いて、上記データベース記憶手段に記憶されたデータベースから検索して、一致する単語データに対応する情報を上記データベース記憶手段から読み出して出力する検索手段とを備えたことを特徴とする情報検索処理装置。

【請求項9】 上記情報検索処理装置はさらに、上記検索手段から出力される情報に基づいて、所定の処理を実行する処理実行手段を備えたことを特徴とする請求項8記載の情報検索処理装置。

【請求項10】 かな文字列を入力する入力手段と、漢字表記の複数の単語データを含む単語辞書を記憶する辞書記憶手段と、

上記辞書記憶手段に記憶された単語辞書を参照して、上記入力手段によって入力されたかな文字列と、上記単語辞書内の単語との間の単語照合を行い、一致したときに漢字表記の単語に変換して単語仮説の文字列として出力する一方、一致しないときにかな文字のまま単語仮説の文字列として出力する単語照合手段と、

上記単語照合手段から出力される単語仮説の文字列に対して、請求項1乃至3、5及び6のうちのいずれか1つに記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して尤度を演算して、演算された尤度に基づいて上記単語仮説の文字列の絞り込みを行った後、絞り込み後の単語仮説の文字列をかな漢字変換後の文字列として出力する絞込手段とを備えたことを特徴とするかな漢字変換装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、学習データ及び学習用テキストデータに基づいて統計的言語モデルを生成する統計的言語モデル生成装置、上記統計的言語モデルを用いて、入力される発声音声文の音声信号を音声認識する音声認識装置、上記音声認識装置を用いた情報検索処理装置及びかな漢字変換装置に関する。

【0002】

【従来の技術】近年、音声認識技術の進展に伴い、音声認識の大語彙タスクへの適用が盛んに行われている。しかしながら、大語彙音声認識のパラダイム（特定領域や時代の支配的な科学的対象把握の方法をいう。）においても、未登録語の問題が完全に解決するわけではない。特に、人名などの固有名詞に関しては、すべてを網羅することが困難であるといった本質的な問題もある。一方で、固有名詞にはタスク達成上重要な情報であるものも多く含まれ、音声認識の実タスク上での運用を考える際、固有名詞の未登録語処理技術は重要な課題となる。

【0003】従来、連続音声認識装置における音素並び（読み）を含めた未登録語の検出方式としては、以下の方法が提案されている。

（１）音素タイプライタ等のサブワードデコーダを併用する方法（以下、第１の従来例の方法という。）、及び（２）サブワードを擬似的な単語として言語モデルに組み込む方法（以下、第２の従来例の方法という。）。

【0004】

【発明が解決しようとする課題】しかしながら、第１の従来例の方法は、別のデコーダを駆動する必要があるため、処理量の観点で望ましくない。また、推定未知語区間の音響スコアには最尤音素系列のスコアが使われるため、語彙内単語系列仮説との統合には、ペナルティやしきい値などのヒューリスティックス（発見的方法）が絡む。

【0005】一方、第２の従来例の方法は、デコーダの変更なしに実現できる利点がある。しかしながら、サブワード系列として得られる未登録語に対し有効な言語処理を行うためには、後処理として、認識語彙よりも大きな語彙による形態素解析などを要する。また、単語とサブワード、あるいは、サブワード間のN-gram確率で、言語的特質を十分反映するモデル化ができるとは考えにくく、認識制約としての有効性に疑問が残る。

【0006】また、電話機における音声認識及び自動ダイヤリング機能や、カーナビゲーションなどの小規模の情報検索装置において、辞書登録の数が限定されるため、対象となる固有名詞の数が限定される。このような場合において、対象の固有名詞が音声認識装置とは別のシステムで管理されるとき、音声認識装置への登録はできず、音声認識率を向上させることができない。

【0007】本発明の目的は以上の問題点を解決し、単語辞書において未登録の未登録語に関する音声認識の精

度を従来例に比較して高くすることができ、未登録語の区間やクラスを同定する統計的言語モデルを生成することができる統計的言語モデル生成装置及び、統計的言語モデル生成装置を用いた音声認識装置を提供することにある。

【0008】また、本発明の別の目的は、電話機における音声認識及び自動ダイヤリング機能や、カーナビゲーションなどの小規模の情報検索処理装置において、単語辞書において未登録の未登録語に関する音声認識の精度を従来例に比較して高くすることができる音声認識装置を用いて情報検索を実行することができる情報検索処理装置を提供することにある。

【0009】さらに、本発明のさらなる目的は、上記統計的言語モデルを利用して、かな漢字変換率を従来技術に比較して向上できるかな漢字変換装置を提供することにある。

【0010】

【課題を解決するための手段】本発明に係る請求項１記載の統計的言語モデル生成装置は、クラス別に分類された単語リストを含む学習データメモリと、その学習データに基づいて単語の読みに対応するサブワード系列の生起確率をクラスに依存して抽出する手段と、その抽出された単語の読みに対応するサブワード系列の生起確率をクラスに依存して評価する手段と、単語とクラスの並びの生起確率を評価する手段と、上記クラスに依存して評価されたサブワード系列の生起確率と上記評価された単語とクラスの並びの生起確率とによって未登録語を含む単語系列の生起確率をサブワードと単語の系列の生起確率として評価する手段と、を備えたことを特徴とする。

【0011】また、請求項２記載の統計的言語モデル生成装置は、請求項１記載の統計的言語モデル生成装置において、上記サブワード系列の生起確率は、長さに関する確率と並びに関する確率とを含むことを特徴とする。

【0012】さらに、請求項３記載の統計的言語モデル生成装置は、請求項１又は２記載の統計的言語モデル生成装置において、上記サブワードとしては、クラスに依存して抽出する単位を用いることを特徴とする。

【0013】本発明に係る請求項４記載の音声認識装置は、入力される発聲音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項１、２又は３記載の統計的言語モデル生成装置によって生成された統計的言語モデルを用いて音声認識することを特徴とする。

【0014】本発明に係る請求項５記載の統計的言語モデル生成装置は、固有名詞又は外来語の普通名詞の単語リストを含む学習データを格納する学習データ記憶手段と、上記学習データ記憶手段に格納された学習データに基づいて、上記学習データにおけるモーラ長に対する単語数の割合が実質的にガンマ分布に従うと仮定したとき

のモーラ長のガンマ分布のパラメータをクラスに依存して推定して計算するとともに、モーラ又はモーラ連鎖であるサブワード単位で、上記固有名詞又は外来語の普通名詞の下位クラスであるクラスを有する第1のN-gramの出現確率を計算することにより未登録語をモデル化したサブワード単位N-gramモデルを生成する第1の生成手段と、所定のテキストデータベースに基づいて生成された単語クラスN-gramモデルと、上記第1の生成手段によって生成されたサブワード単位N-gramモデルと、上記第1の生成手段によって計算されたモーラ長のガンマ分布のパラメータとに基づいて、上記単語クラスと、上記固有名詞又は外来語の普通名詞の下位クラスであるクラスとに依存した第2のN-gramの出現確率を計算することによりサブワード単位に基づいた未登録語を含む統計的言語モデルを生成する第2の生成手段とを備えたことを特徴とする。

【0015】また、請求項6記載の統計的言語モデル生成装置は、請求項5記載の統計的言語モデル生成装置において、上記第1の生成手段によって生成されたサブワード単位N-gramモデルに基づいて、上記サブワード単位を抽出し、抽出したラベルを上記サブワード単位に付与することにより、サブワード単位当たり複数のラベル付きサブワード単位のデータを生成する第3の生成手段と、上記テキストデータベースから抽出された単語と、上記第3の生成手段によって生成された複数のラベル付きサブワード単位のデータとに対して音素並びを付与することにより単語辞書を生成する第4の生成手段とをさらに備えたことを特徴とする。

【0016】またさらに、本発明に係る請求項7記載の音声認識装置は、入力される発声音声文の音声信号に基づいて、所定の統計的言語モデルを用いて音声認識する音声認識手段を備えた音声認識装置において、上記音声認識手段は、請求項5又は6記載の統計的言語モデル生成装置によって生成された統計的言語モデルと、請求項6記載の第4の生成手段によって生成された単語辞書とを用いて音声認識することを特徴とする。

【0017】また、本発明に係る請求項8記載の情報検索処理装置は、上記単語リストに対応する普通名詞の単語データとそれに対応する情報とを含むデータベースを記憶するデータベース記憶手段と、請求項4又は7記載の音声認識装置から出力される音声認識結果の文字列をキーとして用いて、上記データベース記憶手段に記憶されたデータベースから検索して、一致する単語データに対応する情報を上記データベース記憶手段から読み出して出力する検索手段とを備えたことを特徴とする。

【0018】さらに、請求項9記載の情報検索処理装置は、請求項8記載の情報検索処理装置において、さらに、上記検索手段から出力される情報に基づいて、所定の処理を実行する処理実行手段を備えたことを特徴とする。

【0019】またさらに、本発明に係る請求項10記載のかな漢字変換装置は、かな文字列を入力する入力手段と、漢字表記の複数の単語データを含む単語辞書を記憶する辞書記憶手段と、上記辞書記憶手段に記憶された単語辞書を参照して、上記入力手段によって入力されたかな文字列と、上記単語辞書内の単語との間の単語照合を行い、一致したときに漢字表記の単語に変換して単語仮説の文字列として出力する一方、一致しないときにかな文字のまま単語仮説の文字列として出力する単語照合手段と、上記単語照合手段から出力される単語仮説の文字列に対して、請求項1乃至3、5及び6のうちのいずれか1つに記載の統計的言語モデル生成装置によって生成された統計的言語モデルを参照して尤度を演算して、演算された尤度に基づいて上記単語仮説の文字列の絞り込みを行った後、絞り込み後の単語仮説の文字列をかな漢字変換後の文字列として出力する絞込手段とを備えたことを特徴とする。

【0020】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

【0021】＜第1の実施形態＞図1は、本発明に係る第1の実施形態である連続音声認識システムのブロック図である。本発明に係る第1の実施形態の連続音声認識システムは、未登録語モデル生成部20と、サブワード単位データ生成部21と、単語辞書生成部22と、単語クラスN-gramモデル生成部23と、言語モデル生成部24とを備えた統計的言語モデル生成装置を備えたことを特徴としている。

【0022】本実施形態では、未登録語を含む音声の高精度な認識を可能とする、新しい統計的言語モデルを生成する統計的言語モデル生成装置を開示する。本実施形態の統計的言語モデルは、(1)学習データメモリ30内の学習データに基づいて未登録語モデル生成部20によって生成された、未登録語認識用の複数の未登録語モデルである、サブワード単位N-gramモデル及びモーラ長ガンマ分布データと、(2)テキストデータメモリ31内のテキストデータに基づいて単語クラスN-gramモデル生成部23によって生成された単語クラスN-gramモデルと、に基づいて言語モデル生成部24によって統計的言語モデルが生成される。

【0023】これらの未登録語モデルは、各語彙クラスに依存して構築される。ここで、サブワードとは、単語よりも小さい単位をいい、具体的にはモーラ又はモーラ連鎖をいう。モーラとは、韻律論において、強勢や抑揚などの単位となる音の相対的な長さをいい、1モーラは短母音を含む1音節の長さに相当する。日本語では、ほぼ「かな」1字(拗音では2字)がこれに相当する。以下では、固有名詞の下位クラスである、日本人姓及び名

の未登録語を対象を限定して説明する。

0万の日本人の姓のモーラ並び（読み）のデータを含む日本人姓ファイル30aと、(2)表2に示すように約30万の日本人の名のモーラ並び（読み）のデータを含む日本人名ファイル30bとを含む学習データを学習データメモリ30に格納した。

【0025】

【表1】

ス、ズ、キ

ヨ、オ、コ
ト、モ、エ
ト、モ、コ
ケ、イ、コ
.....

* タ、カ、ハ、シ
サ、イ、ト、オ
タ、ナ、カ
ヒ、ラ、ツ、ジ
ア、サ、ギ、ノ
.....

【0026】

【表2】

*

【0027】上記学習データに基づく、本発明者による日本人姓及び名データの分析及び分析結果について説明する。日本人の姓や名をサブワードの系列として眺めるとき、次の特徴を有することが容易に予想される。

(1)長さに関する特徴：姓ではスズキ、サトウ、タカハシなど、名ではヒロシ、アキラ、イチロウなど、3ないし4モーラ長の名前が一般的である。

(2)音素並びに関する特徴：日本人の姓及び名は、基本的に漢字で構成されており、姓ではヤマ、ムラ、ナカなど、名ではロウ、イチ、ヒロなど、高頻度の単位が存在する。

【0028】本発明者は、こうした観点から、日本人姓※モデルの学習データ

※及び名の読みに関する統計的特徴を分析した。人名データとしては約30万人の著名人の名前を集録した公知の人名リストを用いた。この学習データから、漢字と平仮名のみで構成される姓及び名を日本人名として抽出し、得られた姓303,552人分、名295,148人分を対象に分析を行い、その結果を表3に示す。併せて比較のため、日本人姓及び名以外の単語の特徴を分析する。比較する学習データとしては、本特許出願人が所有する自然発話旅行会話データベースより、日本人姓及び名を除いた、のべ1,155,183単語を用いた。

【0029】

【表3】

| | 日本人名 | | 旅行会話 |
|-------|---------|---------|-----------|
| | 姓 | 名 | |
| 単語総数 | 303,552 | 295,148 | 1,161,576 |
| 異なり語彙 | 19,018 | 20,413 | 13,453 |

(注)日本人名の異なり語彙は、音素並び又は読みの異なり単語で評価し、漢字表記の違いは無視した

【0030】本発明者が分析した単語の長さに関する統計を図6に示す。長さの単位としては、モーラ数を用いた。この結果から、日本人姓及び名の長さが3、4モーラを中心に非常に偏った分布を持つことが確認できる。3、4モーラを合わせると、姓及び名ともにほぼ9割の人名が該当することになる。次に、モーラの並びに関する★モーラ並びの偏り

★る統計を表4に示す。モーラ並びの偏りの指標として、頻度上位N種類のモーラ二連鎖による、モーラ並びの被覆率を調べた。ここで、被覆率とは、すべてのモーラの中での二連鎖の占める割合をいう。

【0031】

【表4】

| 二連鎖モーラの種類 (頻度上位N種類) | モーラ並びの被覆率(%) | |
|------------------------|--------------|---|
| | 日本人名 | |
| | 姓 | 名 |
| | 旅行会話 | |

| 9 | 10 |
|------|------|
| 1 | 3.8 |
| 10 | 23.3 |
| 100 | 59.8 |
| 1000 | 84.3 |
| | 4.9 |
| | 28.3 |
| | 66.6 |
| | 82.4 |
| | 0.1 |
| | 5.1 |
| | 19.4 |
| | 35.6 |

(注) 頻度上位N種類の二連鎖モーラによる。モーラ並びの被覆率(%)。奇数

長の単語があるため、被覆率が100%になることはない。

【0032】例えば、日本人姓及び名では、それぞれの高頻度1000種類のモーラ二連鎖だけで、姓及び名におけるモーラ並びの8割以上が被覆される。

【0033】次いで、日本人姓及び名の未登録語モデルに基づく統計的言語モデルの生成方法について詳述する。上述で得られた知見に基づき、日本人姓及び名クラスの未登録語モデルに基づく統計的言語モデルを構築する。また、デコーディングの観点から、統計的言語モデルは、近年広く用いられているN-gram形式で取り扱えることが望ましく、本実施形態では、本未登録語モデルを単語N-gram形式で実装する。

【0034】まず、未登録語を含む単語系列のモデル化について説明する。本実施形態に係る統計的言語モデルのベースとなるのは、単語のクラスN-gramモデルである。この単語クラスN-gramモデルは、単語クラスN-gramモデル生成部23により、テキストデータベースメモリ31に格納された多数の日本語の書き下し文からなるコーパスと呼ばれるテキストデータに基づいて、例えば公知の最尤推定法を用いて単語のクラスタリングを行って単語クラスN-gramモデルを生成し、このモデルを単語クラスN-gramモデルメモリ43に格納する。

【0035】単語クラスN-gramモデルでは、単語系列 $W = w_1, w_2, w_3, w_4, \dots, w_T$ の言語的尤度 $p_h(W)$ が一般に次式で与えられる。ただし、 w_t は単語系列 W の t 番目の単語であり、 c^* は単語 w_t の語彙クラスを表わすものとする。

【0036】

【数1】

$$p_h(W) = \prod_t p(w_t | c^{w_t}) \cdot p(c^{w_t} | c^{w_{t-n+1}}, \dots, c^{w_{t-1}})$$

【0037】ところで、単語 w_t (以下、1つの単語を w で表す。)には認識語彙にない未登録語が含まれている。これら未登録語の生起確率を音素並び (又は読み) の統計的特徴に基づいて推定するとき、上記数1中のクラス内単語1-gram確率 $p(w | c^*)$ は次式により与えられる。ただし、 M^* は単語 w のモーラ系列を表す。

【0038】

【数2】(1) もし $w \in$ 単語辞書のとき

$$p(w | c^*) \equiv (1 - p(OOV | c^*)) \cdot p(w | c^*, \text{inVoc}),$$

(2) もし上記以外のとき

$$p(w | c^*) \equiv p(OOV | c^*) \cdot p(M^* | c^*, OOV)$$

10 【0039】ここで、単語辞書は、単語辞書メモリ12に格納される語彙辞書であり、OOVは未登録語であり、inVocは単語辞書内の条件を表す。上記数2において、確率 $p(OOV | c^*)$ は、クラスから未登録語が生起する確率であり、例えば公知の方法 (例えば、従来技術文献「政龍ほか、"品詞及び可変長形態素列の複合N-gramを用いた形態素解析"、言語処理学会誌「自然言語処理」、Vol. 6, No. 2, pp. 41-57, 1999年」など参照。)で推定できる。この推定方法について説明すると、公知のチューリング(Turing)推定法を用いたとき、データ上に r 回出現する形態素は、次式の r^* 回と推定される。

【0040】

$$\text{【数3】 } r^* = \{(r+1)n_{r+1}\} / n_r$$

【0041】ここで、 n_r はデータ上に r 回出現した形態素の種類数を表す。従って、 r 回出現する形態素 w の品詞からの出現確率 $P(w | c_e)$ は、次式で表される。

【0042】

$$\text{【数4】 } P(w | c_e) = r^* / N(c_e)$$

30 【0043】上記出現確率 $P(w | c_e)$ を、クラス c_e に属する全ての形態素について計算し、次式に示すように、1から引いた残りが品詞 c_e から未知語出現する確率 $P(c_{he})$ である。

【0044】

【数5】

$$P(c_{he}) = 1 - \sum_{w \in c_e} P(w | c_e)$$

40 【0045】本実施形態では、限られた評価セット上で未登録語モデルの有効な評価を行うことに主眼を置き、次式を用いてモデル化を行う。すなわち、未登録語の生起は予め規定したいいくつかのクラス (ここで、クラスの集合を C^{OOV} という。)のみに許すこととし、これらクラスからの単語生起は全て未登録語モデルで説明することとし、登録語を作らない。

【0046】

【数6】

$$\begin{aligned} p(OOV | c^w) \\ = 0, \text{もし } c^w \notin C^{OOV} \text{ のとき} \\ = 1, \text{もし } c^w \in C^{OOV} \text{ のとき} \end{aligned}$$

【数7】

$$\begin{aligned}
 & p(w | c^w) \\
 &= p(w | c^w, \text{inVoc}), \text{もし } c^w \in C^{\text{OOV}} \text{ のとき} \\
 &= p(M^w | c^w, \text{OOV}) = p(M^w | C^w), \text{もし } c^w \in C^{\text{OOV}} \text{ のとき}
 \end{aligned}$$

【0047】次いで、日本人姓及び名の未登録語モデルについて説明する。上述したように、日本人姓及び名の読みには、モーラ長、及びモーラ並び、それぞれに関して特徴的な傾向が見られた。従って、上記数7の未登録語モデル $p(M^w | c^w)$ は、次式のように展開することにより、高精度なモデル化が可能となる。ただし、 $\text{len}(M^w)$ は単語 w のモーラ長を表す。

【0048】

$$p(M^w | c^w) = p(\text{len}(M^w) | c^w) \cdot p(M^w | c^w, \text{len}(M^w))$$

【0049】上記数8の確率 $p(\text{len}(M) | c)$ は、日本人姓又は名クラスにおいて、長さ $\text{len}(M)$ の単語が生起する確率である。本実施形態では、その確率分布が次式で与えられるガンマ分布に実質的に従うことを仮定する。すなわち、上記数8の右辺は、第1項のモーラ長のガンマ分布の確率と、第2項のサブワード単位バイグラムの確率との掛け算になっている。ただし、 α 、 λ はクラス c に依存するパラメータであり、モーラ長の平均と分散より定まる。

【0050】

【数9】

$$g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x) \quad (x \geq 0)$$

ここで、

【数10】

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx$$

【0051】一方、上記数8の確率 $p(M^w | c^w, \text{len}(M^w))$ は、クラス c^w において長さ $\text{len}(M^w)$ のモーラ並びが $M^w = m_1^w, m_2^w, \dots$ となる確率であり、次式のサブワード単位 $N\text{-gram}$ によりモデル化する。ただし、 $U = u_1, u_2, \dots$ は詳細後述する手法で自動獲得したサブワード単位（モーラ又はモーラ連鎖）の系列である。また、式中のサブワード単位 $N\text{-gram}$ には終端記号への遷移を含まない。

【0052】

【数11】

$$\begin{aligned}
 & p(M^w | c^w, \text{len}(M^w)) \\
 &= \prod_{i=1}^{\text{len}(M^w)} p(m_i^w | m_1^w, \dots, m_{i-1}^w, c^w) \\
 &= \sum_{U, \text{ s. t. } U \text{ eq } M^w} \prod p(u_i | u_1, \dots, u_{i-1}, c^w) \\
 &\equiv \max_{U, \text{ s. t. } U \text{ eq } M^w} \prod p(u_i | u_1, \dots, u_{i-1}, c^w) \\
 &\equiv \max_{U, \text{ s. t. } U \text{ eq } M^w} \prod p(u_i | u_{j-n+1}, \dots, u_{j-1}, c^w)
 \end{aligned}$$

【0053】以上述べてきた本実施形態に係る統計的言語モデルにおいて、「... あさぎ野 陽子 と ...」が出力される例を図7に示す。例では、日本人姓及び名クラスの単語「あさぎ野」、「陽子」に対して、クラスラベル付きモーラ系列「アサギノ（日姓）」、「ヨオコ（日名）」が出力される。本モデルでは、日本人姓及び名の生起に対して、次の3レベルから言語的制約をかける。

【0054】<3レベルの言語的制約>

(1) 単語間制約：単語のクラス $N\text{-gram}$ を用い、単語コンテキストにおいて日本人姓及び名（クラス）が生起する尤度を評価する。サブワードによる姓及び名のモデル化は下位の階層に隠蔽されるため、登録語系列のモデル化には悪影響を及ぼさない。

(2) 姓及び名区間の継続長制約：姓及び名それぞれのモーラ長に関するガンマ分布を用い、区間の姓及び名らしさを評価する。この制約により、不当に短い、もしくは長いモーラ系列の湧き出しを防ぐことができる。

(3) サブワードの並び制約：モーラとモーラ連鎖を単位とする姓及び名のサブワード単位 $N\text{-gram}$ を用いる。モーラ連鎖を単位とすることで、 $N\text{-gram}$ の高精度化が期待できる。ここで、モデル化単位とするモーラ連鎖は、後述する繰り返し学習において自動的に獲得する。

【0055】次いで、未登録語モデル生成部20によって実行される未登録語モデルの学習生成処理について説明する。未登録語モデル生成部20は、学習データメモリ30に格納された日本人姓ファイル30a及び日本人名ファイル30bに基づいて、日本人の姓クラス（ラベルでは、日姓と略記する。）及び日本人の名クラス（ラベルでは、日名と略記する。）の未登録語モデルを構築する。この学習生成処理では、具体的には、サブワード単位 $N\text{-gram}$ モデルとモーラ長ガンマ分布のデータを生成する。以下の実施形態では、個人名はそれぞれ等しい確率で出現するとし、各姓又は各名の観測頻度として人名リスト中の同姓者又は同名者の人数を用いること

とする。サブワード単位N-gramには、初期単位セットとして単一モーラのみを与え、後述の繰り返し学習において、逐次的に新たなモーラ連鎖を単位セットに追加していく。これら単位候補となるモーラ連鎖には頻度による予備選択を施すことで、学習の効率化を図った。

【0056】図3は、図1の未登録語モデル生成部20によって実行される未登録語モデル生成処理を示すフローチャートであり、図4は、図3のサブルーチンであるサブワード2-gramの単位決定処理（ステップS4）を示すフローチャートである。

【0057】未登録語モデルの学習データとして、モーラ系列で表現された姓及び名のリストを用い、このリストデータはそれぞれ学習データメモリ30内の日本人姓ファイル30a及び日本人名ファイル30bに格納されている。この学習データは先の表1及び表2の通りであって、各モーラをカンマ「，」で区切って表記してある。この学習データに基づいて、未登録語モデル、すなわち、モーラ長のガンマ分布とサブワード単位N-gramモデルを生成する。以下では、N-gramの次数Nが2の場合について、未登録語モデルの生成手順を説明する。

【0058】図3のステップS1において学習データメモリ30から学習データを読み出し、ステップS2において、読み出した学習データに基づいて、姓又は名の1単語当たりのモーラ数の平均 μ と分散 V を計算した後、次式を用いてモーラ長のガンマ分布のパラメータを推定する。

【0059】

【数12】 $\lambda = V / \mu$

【数13】 $\alpha = \mu^2 / V$

【0060】さらに、ステップS3において上記学習データに基づいて、サブワード単位2-gramの単位候補となる、高頻度のモーラ連鎖を抽出し、抽出した単位候補を「単位候補セット」と呼ぶ。ここでは、学習データ上に出現する、例えば長さ2以上の全てのモーラ連鎖に対して、その頻度を調べ、頻度が所定の値（＝100）以上のモーラ連鎖を単位候補として抽出する。次いで、ステップS4において、図4に示すサブルーチンであるサブワード2-gramの単位決定処理を実行し、最後に、ステップS5において、学習終了後の暫定未登録語モデルをサブワード単位N-gramモデルとしてサブワード単位モデルN-gramモデルメモリ40に格納するとともに、モーラ長のガンマ分布のデータをモーラ長ガンマ分布データメモリ41に格納する。

【0061】図4のサブワード2-gramの単位決定処理においては、まず、ステップS11において確定単位セットにすべての単一のモーラを挿入し、ステップS12において単位候補セットの中から1つのモーラ連鎖を選択する。次いで、ステップS13において選択したモーラ連鎖は確定単位セットに含まれているか否かを判

断し、YESのときはステップS15に進む一方、NOのときは、ステップS14に進む。ステップS14において選択したモーラ連鎖を確定単位セットに追加して、公知の最尤推定法を用いて暫定サブワード2-gramモデルを生成してステップS15に進む。ここで、2-gramモデルは、学習データと追加セットとの間の1-gramと、追加セットのみの0-gramを用いて、公知の削除補間法（例えば、従来技術文献「中川聖一，「確率モデルによる音声認識」，社団法人電子情報通信学会編，pp. 63-64，昭和63年7月1日発行」など参照。）を用いて補間して生成する。この暫定サブワード2-gramと、モーラ長のガンマ分布データを合わせて「暫定未登録語モデル」と呼ぶ。ステップS15においては、すべてのモーラ連鎖について上記ステップS13及びS14の処理したか否かを判断し、NOであるときはステップS12に戻り上記の処理を繰り返すが、YESのときはステップS16に進む。ステップS16において各暫定未登録語モデルに対して数8を用いて平均尤度を計算し、平均尤度を最大にする暫定未登録語モデルの単位セットを新しい確定単位セットとする。そして、ステップS17において確定単位セットに含まれるモーラ連鎖の数 \geq 所定のしきい値 N_{th} （例えば、150である。）であるか否かが判断され、NOのときはステップS12に戻り上記の処理を繰り返す一方、YESのときは元のメインルーチンに戻る。

【0062】図9は、図1の未登録語モデル生成部20によって実行される未登録語モデル生成処理における、モーラ連鎖の単位化による平均尤度の向上を示すグラフであって、モーラ連鎖の種類の数に対する平均尤度を示すグラフである。すなわち、図9は繰り返し学習における平均尤度（数8）の変化を示す。単位候補とするモーラ連鎖は、頻度が100以上のものとした。表1乃至表3に示す学習データからは、姓モデルで1,829種類、名モデルで1,660種類のモーラ連鎖が単位候補となる。サブワード単位N-gramは $N=2$ とし、1-gram、0-gramを用いた公知の削除補間法で補間した。図9に示すように、モーラ連鎖をサブワード単位として追加していくことで、学習データに対する平均尤度は単調に上昇する。モーラ連鎖を150個追加したモデルの平均尤度は、モーラ連鎖を用いないモデルに比べ、姓モデルで3.9倍、名モデルで3.2倍となった。未登録語モデルを単語1-gramとみなすと、単語の学習セットパーブレキシティは姓モデルで74%、名モデルで69%改善されることになる。

【0063】次いで、未登録語モデルを単語N-gram形式による単語辞書に実装する方法について説明する。上述の未登録語モデルは、以下に述べる方法により、近似なく、クラスN-gramの形式で扱うことができる。そのため、統計的言語モデルとしてクラスN-gramを扱うことが可能なデコーダであれば、デコー

ダの変更無しに、本方法による未登録語の認識が可能となる。ただし、極端に長い未登録語（本実施形態では、10モーラ以上の姓及び名）が認識対象とならないことが条件となる。サブワード単位N-gramで単位として用いるモーラ及びモーラ連鎖は、擬似的な単語として扱い、認識辞書、及びクラスN-gramに組み込む。その際、各サブワード単位は以下のラベル付けによる展開を行い、ラベル違いの同一サブワード単位を複数生成する。

【0064】すなわち、ラベル付きサブワード単位データ生成部20は、サブワード単位N-gramメモリ42に格納された、日本人姓及び名のサブワード単位N-gramそれぞれに対して以下の処理を実行し、この結果得られるラベル付きサブワード単位群のデータをラベル付きサブワード単位データメモリ40に格納する。すなわち、日本人姓（もしくは日本人名）サブワード単位N-gramで単位として用いられるサブワード単位（単一モーラ、モーラ連鎖）を全て抽出する。次いで、抽出したそれぞれのサブワード単位に依りて、ラベルを複数生成する。生成した各ラベルを当該サブワード単位に付与することにより、サブワード単位あたり複数のラベル付きサブワード単位が生成される。ここでラベルとは、(a) クラス（本実施形態においては、「日姓」又は「日名」）、(b) 単語内の開始位置（1, 2, ..., LenMax+1；サブワード単位のモーラ数：ここで、LenMaxは事前に設定された値であり、認識対象とする最長の姓もしくは名のモーラ数を意味する。）、及び(c) 単語の終端であるか否か（終端を示す「終」又は「-」）の3項組みである。

【0065】上記(b)の開始モーラ位置による展開は、学習データに出現する最長の姓及び名に合わせ、ともに終端位置が9モーラまでとなるようにした。上記(c)で単語終端ラベルを付与したサブワード単位には、音素並び（読み）の終端にポーズが入ることを許容する。ここで、ラベル付きサブワードは、その遷移に次の制約を受ける。(i) 登録語のクラスからラベル付きサブワードへの遷移は、ラベル付きサブワードの開始モーラ位置が1の場合のみ許される。逆に、(ii) ラベル付きサブワードから登録語のクラスへの遷移は、ラベル付きサブワードに単語終端ラベルが付与されている場合のみ許される。(iii) ラベル付きサブワード間の遷移は、単語内でのモーラ位置が接続し、かつ同一のクラスに属する場合のみ許される。

【0066】さらに、単語辞書生成部22は、ラベル付きサブワード単位データメモリ42内のデータと、テキストデータメモリ31内のテキストデータとに基づいて以下のように単語辞書を生成して単語辞書メモリ12に格納する。まず、単語辞書生成部22は、テキストデータベースメモリ31内のテキストデータ中に出現する全ての単語を抽出し、単語辞書メモリ12に格納する。次

いで、ラベル付きサブワード単位データメモリ42中の全てのラベル付きサブワード単位を単語辞書メモリ12に格納する。単語辞書メモリ12に格納された全てのエントリに対して、入手で、もしくは、読みと音素系列との対応テーブルを用いて音素付与を行う公知の音素付与処理プログラムを用いて、その読み（音素系列）を付与することにより単語辞書を生成する。これによって、単語辞書メモリ12に単語辞書が生成格納される。

【0067】次いで、言語モデル生成部24の統計的言語モデル生成処理について説明する。ラベル付きサブワード単位 u_i を単語として扱う際、クラスN-gram ($N=2$)における確率は、以下のように与える ($N>2$ でも同様)。ただし、 u はラベル付け前のサブワード単位を表し、 $\#$ は単語内のモーラ開始位置記号を表し、 c は語彙クラスを表し、 $len(u)$ はサブワード単位 u のモーラ長、 $p_{i,w}(u_i | u_{i-1}, c)$ はクラスに依存したサブワード単位2-gramであり、 $p_{i,w}(c' | c)$ はクラス2-gramのクラス間遷移確率を表す。また、 $classof(u_i)$ 、 $startof(u_i)$ 、 $isend(u_i)$ はそれぞれ、ラベル付きサブワード単位 u_i のラベルである語彙クラス、単語内のモーラ開始位置、単語終端か否か、を参照する関数であり、 $endof(u_i)$ は $startof(u_i) + len(u)$ により与えるものとする。 $G_c(*)$ はクラスのモーラ長ガンマ分布 $g_c(x)$ （数9）に基づく確率関数であり、次の定積分により与えられる。

【0068】

【数14】

$$G_c(len=x) = \int_{len-0.5}^{len+0.5} g_c(x) dx$$

【数15】

$$G_c(x > len) = \int_{len+0.5}^{\infty} g_c(x) dx$$

【0069】まず、先頭サブワードの生起確率である姓及び名クラスのクラス内1-gram確率 $p(u_i | c)$ は、ラベル付きサブワード単位 u_i の開始モーラ位置ラベルが1である場合のみ許す。従って、次式で表すことができる。

【0070】

【数16】(1) もし $(classof(u_i) = c) \wedge (startof(u_i) = 1)$

$\wedge (isend(u_i) = \text{偽})$ のとき

$$p(u_i | c) = p_{i,w}(u_i | \#, c) \cdot G_c(x > len(u_i))$$

(2) もし $(classof(u_i) = c) \wedge (startof(u_i) = 1)$

$\wedge (isend(u_i) = \text{真})$ のとき

$$p(u_i | c) = p_{i,w}(u_i | \#, c) \cdot G_c(x = len(u_i))$$

(3) もし上記以外るとき

$$p(u_i | c) = 0$$

【0071】次いで、第1のクラス間2-gram確率である、ラベル付きサブワード単位間の遷移は、両者のクラスが同じで、かつ、両者の単語内でのモーラ位置が接続する場合のみ許す。従って、第1のクラス間2-gram確率 $p(u_{d1} | u_{d1})$ は次式で表すことができる。

【0072】

【数17】 (1) もし $(classof(u_{d1}) = classof(u_{d1}) = c)$
 $\wedge (startof(u_{d1}) = endof(u_{d1}) + 1)$
 $\wedge (isend(u_{d1}) = \text{偽})$ のとき
 $p(u_{d1} | u_{d1}) = p_{sw}(u_1 | u_1, c) \cdot G_c(x > (endof(u_{d1}) + len(u_{d1}))) / G_c(x > endof(u_{d1}))$
 (2) もし $(classof(u_{d1}) = classof(u_{d1}) = c)$
 $\wedge (startof(u_{d1}) = endof(u_{d1}) + 1)$
 $\wedge (isend(u_{d1}) = \text{真})$ のとき
 $p(u_{d1} | u_{d1}) = p_{sw}(u_1 | u_1, c) \cdot G_c(x = (endof(u_{d1}) + len(u_{d1}))) / G_c(x > endof(u_{d1}))$
 (3) もし上記以外のとき
 $p(u_{d1} | u_{d1}) = 0$

【0073】さらに、第2のクラス間2-gram確率である、ラベル付きサブワード単位から、次単語のクラスへの遷移は、ラベル付きサブワード単位に単語終端ラベルが付与されている場合にのみ許す。従って、第2のクラス間2-gram確率 $p(c | u_d)$ は次式で表される。

【0074】

【数18】 (1) もし $isend(u_d) = \text{真}$ のとき
 $p(c | u_d) = p_{sw}(c | classof(u_d))$
 (2) もし上記以外のとき
 $p(c | u_d) = 0$

【0075】図5は、図1の言語モデル生成部24によって実行される言語モデル生成処理を示すフローチャートである。図5において、まず、S21において各メモリ40、41、42からそれぞれ格納された各データを40
 読み出し、ステップS22においてモーラ長のガンマ分布のデータに基づいて、数14及び数15を用いて確率関数の値を計算する。次いで、ステップS23において先頭サブワードの生起確率であるクラス内-gram確率を数16を用いて計算し、ステップS24においてサブワード間の遷移確率である第1のクラス間2-gram確率を数17を用いて計算し、ステップS25において終端サブワードから次単語のクラスに遷移する確率である第2のクラス間2-gram確率を数17を用いて
 計算する。そして、ステップS26において上記計算さ

れた確率をまとめて、未登録語モデルに基づいて統計的言語モデルとして統計的言語モデルメモリ44に格納する。

【0076】上記の言語モデル生成処理で生成された統計的言語モデルの別の一例を図8に示す。図8の例における遷移確率は上述で定義されたものである。

【0077】次いで、図1に示す連続音声認識システムの構成及び動作について説明する。図1において、単語照合部4に接続された音素隠れマルコフモデル（以下、隠れマルコフモデルをHMMという。）メモリ11内の音素HMMは、各状態を含んで表され、各状態はそれぞれ以下の情報を有する。

(a) 状態番号、(b) 受理可能なコンテキストクラス、(c) 先行状態、及び後続状態のリスト、(d) 出力確率密度分布のパラメータ、及び(e) 自己遷移確率及び後続状態への遷移確率。なお、本実施形態において用いる音素HMMは、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HMMを変換して生成する。ここで、出力確率密度関数は34次元の対角共分散行列をもつ混合ガウス分布である。また、単語照合部4に接続された単語辞書メモリ12内の単語辞書は、音素HMMメモリ11内の音素HMMの各単語毎にシンボルで表した読みを示すシンボル列を格納する。

【0078】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、 Δ 対数パワー及び16次 Δ ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して単語照合部4に入力される。

【0079】単語照合部4は、ワンパス・ヒタビ復号化法を用いて、バッファメモリ3を介して入力される特徴パラメータのデータに基づいて、音素HMMメモリ11内の音素HMMと、単語辞書メモリ12内の単語辞書とを用いて単語仮説を検出し尤度を計算して出力する。ここで、単語照合部4は、各時刻の各HMMの状態毎に、単語内の尤度と発声開始からの尤度を計算する。尤度は、単語の識別番号、単語の開始時刻、先行単語の違い毎に個別にもつ。また、計算処理量の削減のために、音素HMM及び単語辞書とに基づいて計算される総尤度のうちの低い尤度のグリッド仮説を削減する。単語照合部4は、その結果の単語仮説と尤度の情報を発声開始時刻からの時間情報（具体的には、例えばフレーム番号）とともにバッファメモリ5を介して単語仮説絞込部6に出力する。

【0080】単語仮説絞込部6は、単語照合部4からバッファメモリ5を介して出力される単語仮説に基づいて、統計的言語モデルメモリ44内の統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の

単語の単語仮説に対して、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有する仮説の単語列を認識結果として出力する。なお、タスク適応化された統計的言語モデルは、各タスク毎に1つの統計的言語モデルを備え、単語仮説絞り込み部6は、音声認識しようとするタスクに対応する統計的言語モデルを選択的に参照する。本実施形態においては、好ましくは、処理すべき当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の2つの音素とを含む3つの音素並びをいう。

【0081】例えば、図2に示すように、 $(i-1)$ 番目の単語 W_{i-1} の次に、音素列 a_1, a_2, \dots, a_n からなる i 番目の単語 W_i がくるときに、単語 W_{i-1} の単語仮説として6つの仮説 $W_a, W_b, W_c, W_d, W_e, W_f$ が存在している。ここで、前者3つの単語仮説 W_a, W_b, W_c の最終音素は $/x/$ であるとし、後者3つの単語仮説 W_d, W_e, W_f の最終音素は $/y/$ であるとする。終了時刻 t_e と先頭音素環境が等しい仮説（図2では先頭音素環境が $"x/a_1/a_2"$ である上から3つの単語仮説）のうち総尤度が最も高い仮説（例えば、図2において1番上の仮説）以外を削除する。なお、上から4番めの仮説は先頭音素環境が異なるため、すなわち、先行する単語仮説の最終音素が x ではなく y であるので、上から4番めの仮説を削除しない。すなわち、先行する単語仮説の最終音素毎に1つのみ仮説を残す。図2の例では、最終音素 $/x/$ に対して1つの仮説を残し、最終音素 $/y/$ に対して1つの仮説を残す。

【0082】以上の実施形態においては、固有名詞の下位クラスである日本人の姓及び名とを、未登録語の語彙クラスとして用いているが、本発明はこれに限らず、以下の固有名詞や外来語の普通名詞などに適用することができる。

(1) 外国人の姓と名、(2) 地名、(3) 会社名、(4) 各種施設名、(5) 各種製品名など。従って、本発明では、未登録語の語彙クラスとして、固有名詞及び外来語の普通名詞の各下位クラスを用いることができる。

【0083】また、本実施形態で用いるクラス依存未登録語モデルは、クラス毎に異なったパラメータ構造を持たせることが可能となっている。そのため、各クラスにおける読みの統計的特徴を強く反映させたモデル化が可能である。実施形態において、日本人姓及び名の未登録語モデル構築には、そのパラメータ構造として、(1) 単語長のガンマ分布と、(2) 終端記号への遷移を含まないサブワード単位 $N-gram$ とを用いた。しかしながら、その他のクラス、例えば宿泊施設名（「京都第一

観光ホテル」、「赤坂プリンスホテル」、「いとう旅館」、…)のように、複合語を多く含むクラスをモデル化するには、(1) 長さに関する制約が有効とはならない場合がある。その場合は、上記(1)の制約を省き、代わりに上記(2)の制約に基づいたサブワード単位 $N-gram$ において、終端記号への遷移を含むようモデル化する（例えば、「ホテル」や「旅館」から終端記号へ高い確率で遷移すること、そうしたクラスに対しても高精度な未登録語モデルを構築することが可能である。

【0084】以上の実施形態においては、当該単語の先頭音素環境とは、当該単語より先行する単語仮説の最終音素と、当該単語の単語仮説の最初の2つの音素とを含む3つの音素並びとして定義されているが、本発明はこれに限らず、先行する単語仮説の最終音素と、最終音素と連続する先行する単語仮説の少なくとも1つの音素とを含む先行単語仮説の音素列と、当該単語の単語仮説の最初の音素を含む音素列とを含む音素並びとしてもよい。

【0085】以上の実施形態において、特徴抽出部2と、単語照合部4と、単語仮説絞り込み部6と、未登録語モデル生成部20と、サブワード単位データ生成部21と、単語辞書生成部22と、単語クラス $N-gram$ モデル生成部23と、言語モデル生成部24とは、例えば、デジタル電子計算機などのコンピュータで構成され、バッファメモリ3、5と、音素HMMメモリ11と、単語辞書メモリ12と、学習データメモリ30と、テキストデータベースメモリ31と、サブワード単位 $N-gram$ モデルメモリ40と、モーラ長ガンマ分布データメモリ41と、ラベル付きサブワード単位データメモリ42と、単語クラス $N-gram$ モデルメモリ43と、統計的言語モデルメモリ44とは、例えばハードディスクメモリなどの記憶装置で構成される。

【0086】以上実施形態においては、単語照合部4と単語仮説絞り込み部6とを用いて音声認識を行っているが、本発明はこれに限らず、例えば、音素HMMメモリ11内の音素HMMを参照する音素照合部と、例えばOne Pass DPアルゴリズムを用いて統計的言語モデルを参照して単語の音声認識を行う音声認識部とで構成してもよい。

【0087】

【実施例】本発明者は、本実施形態に係る統計的言語モデルの有効性を確認するため、音声認識実験を行った。以下では、二種類の統計的言語モデルを用いて比較評価を行う。両言語モデルは、共通のベースモデルとして、表3の旅行会話データのみから生成したクラス $N-gram$ を用いる。このベースモデルに対して、それぞれの方法で日本人姓クラス、及び名クラスのクラス内単語 $1-gram$ を置換する。

【0088】評価を行う統計的言語モデルは以下の通り

10

20

30

40

50

である。

(1) 本実施形態に係る統計的言語モデル：日本人姓及び名クラスの単語1-gramとして、姓及び名それぞれの未登録語モデルを用いる。サブワード単位N-gramで単位として用いるモーラ連鎖は、特に断らない限り150個の場合を評価する。認識語彙は、日本人姓及び名以外の単語12,755単語+サブワードで構成し、登録語の日本人姓及び名は作らない。

(2) 登録語方法（以下、比較例という。）：日本人姓及び名クラスの単語1-gramとして、表3の人名データによる単語1-gramを用いる。認識語彙は、日本人姓及び名以外の単語12,755単語+日本人姓及び名39,431単語となる。この方法は、評価セット中のほぼ全人名をカバーする語彙を持つこと、また、本実施形態の方法が未登録語モデルの最尤推定に用いる人名データを単語1-gramとして直接用いることから、概ね本実施形態の方法による認識精度の上限値を与えるものと考えられる。

【0089】これら2つの方法の音声認識率を、以下の基準により評価する。

音声認識率

| 認識率(%) | 本実施形態 | 比較例 |
|-----------|-------|-------|
| 単語認識率 | 87.51 | 87.30 |
| 姓及び名单語再現率 | 70 | 73 |
| 適合率 | 67 | 75 |

(注) 姓及び名は、読み、クラス、区間が全て正しい場合のみ正解として評価した。比較例の方法の認識率は、概ね本実施形態の方法の上限値に相当すると考えられる。

【0093】本実施形態の方法では、未登録語である姓及び名を、登録語として認識した場合とほぼ同等の精度で認識できた。予想に反し、本実施形態の方法の単語認識率が比較例の方法を上回った理由の一つとして、以下が挙げられる。音響尤度の低い一部の姓及び名に対し、本実施形態の方法では読み誤りはあるものの区間が正しく検出され、結果、前後の単語にまで認識誤りを誘発す※
姓及び名单語の区間検出率

| 認識率(%) | 本実施形態 | 比較例 |
|-----------|-------|-----|
| 姓及び名区間再現率 | 87 | 80 |
| 適合率 | 84 | 82 |

(注) 姓及び名のクラスや区間が正しい場合を、正解として評価（音響尤度の影響が強い読み誤りは無視する。）した。

【0095】図10は、本発明者による実験の実験結果であって、日本人の姓及び名の再現率におけるモーラ連鎖の単位化効果を示すグラフであり、モーラ連鎖の種類

* (1) 単語認識率：評価データに出現する全単語の認識率を評価する。日本人姓及び名は、クラス（「日姓」又は「日名」）、読み（モーラ並び）、位置（DPによる対応付け）が全て正しい場合のみを正解とする。ただし、読みに関し、明らかに等価な長音（ヨウコとヨオコ）は手作業で修正して評価した。

(2) 姓及び名单語の再現率及び適合率：単語認識率評価時の動的計画法のマッチング（DPマッチング）を用いて、日本人姓及び名のみの再現率と適合率を評価する。

【0090】ここで、評価セットには、旅行会話ドメインの42片側会話4,990単語を用いた。評価セットに出現する日本人名は、姓及び名、合わせて70単語（異なり単語数52）である。うち、表3の人名リストにも出現しない姓は3単語（アサギノ1単語、チンザイ2単語）であった。

【0091】次いで、表3に本実施形態の方法、及び比較例の方法の音声認識率を示す。

【0092】

*20 【表5】

※ることが少なかったと考えられる。このことは、次の表6に示す読み誤りを無視した姓及び名区間の再現率及び適合率において、本実施形態の方法が優れていることから裏付けられる。

【0094】

【表6】

の数に対する単語再現率を示すグラフである。図10から明らかなように、単位化するモーラ連鎖を増やすことで、モデルによる姓及び名の尤度が上がり、再現率が改

善されるものと思われる。これは、上述の学習セットに対する平均尤度の改善傾向と合致する。

【0096】次いで、希有な姓及び名に対する音声認識率について説明する。本実施形態で提案する未登録語モデルの利点は、事前に予測できない希有な単語も正しく認識できる可能性があることにある。ここでは、そうした希な姓及び名を模擬的に作り出すことで、本実施形態の方法の評価を行う。評価セットには、52種類の日本人姓及び名が出現する。そこで、これらの単語と同じ読みを持つ全ての姓及び名を表1の学習データから削除した後、前節と同様に本実施形態の方法と比較例の方法による音声認識率の比較実験を行った。表7にその結果を示す。

【0097】

【表7】希有な姓及び名入力時の音声認識率

| 認識率(%) | 本実施形態 | 比較例 |
|-----------|-------|-------|
| 単語認識率 | 86.66 | 86.08 |
| 姓及び名单語再現率 | 31 | 6 |
| 適合率 | 36 | 8 |

(注) 学習に用いる姓及び名データから、評価セットに出現する姓及び名と同じ読みを持つエントリを全て削除して実験。姓及び名は、読み、クラス、区間が全て正しい場合のみ正解として評価した。登録語方式の再現率・適合率が0%にならないのは、形態素の不備により、一部の姓が「普通名詞」になっていたためである。

【0098】表5から明らかなように、本実施形態の方法では、学習データに存在しない姓及び名を与えても、31%の再現率で、その読み、クラス、区間を正しく認識できた。結果、単語認識率でも登録語方式を0.58ポイント上回った。

【0099】以上説明したように、本発明に係る実施形態によれば、未登録語モデルのクラス依存化により、次の特有の効果を得ることができる。

(1) モデル化対象を限定することで、読みの統計的特徴をより明確化することができ、クラス固有のパラメータ制約を導入できるため、未登録語モデルを高精度化することができる。

(2) 検出区間の言語処理が可能である。未登録語は、読みに加えクラスも同時に同定される。読みとクラスは、固有名詞の言語処理において必要十分な情報となるケースが多いものと考えられる。

(3) 上記生成された統計的言語モデルを用いて音声認識することにより、従来技術に比較して高い認識率で音声認識することができる。

【0100】＜第2の実施形態＞図11は、本発明に係る第2の実施形態である連続音声認識システムの構成を示すブロック図であり、図12は、図11の連続音声認

識システムを用いた、自動ダイヤリング機能付き電話機の構成を示すブロック図である。

【0101】図11の連続音声認識システムは、図1の連続音声認識システムに比較して以下の点が異なる。

(1) 学習データメモリ30は、日本人姓ファイル30aと、日本人名ファイル30bとに加えて、地名ファイル30cなどのファイルを含む。地名ファイルは、例えば日本や外国の地名の単語を含むファイルである。

(2) 図1の単語クラスN-gramモデル生成部23に代えて、有限状態オートマトンモデル生成部23aを備える。

(3) 図1の単語クラスN-gramモデルメモリ43に代えて、有限状態オートマトンモデルメモリ43aを備える。

【0102】ここで、有限状態オートマトンモデル生成部23aは、テキストデータベースメモリ31に格納された多数の日本語の書き下し文からなるコーパスと呼ばれるテキストデータに基づいて有限状態オートマトンモデルを生成し、このモデルを有限状態オートマトンメモリ43aに格納する。

【0103】図11の連続音声認識システムにおいて、破線で囲んだ部分を音声認識装置100という。すなわち、音声認識装置100は、マイクロホン1から単語仮説較込部6までの回路及び処理部と、単語照合部4に接続された音素HMMメモリ11及び単語辞書メモリ12、並びに、単語仮説較込部6に接続された統計的言語モデルメモリ44とを備えて構成される。第2の実施形態では、この音声認識装置100を用いて、図12の自動ダイヤリング機能付き電話機が構成される。なお、単語辞書メモリ12及び統計的言語モデルメモリ44内のデータは予め図11のシステムにより生成されて格納される。従って、音声認識装置100は、マイクロホン1に入力される人名などの単語の話者音声にตอบสนองして、音声認識処理を実行して、音声認識結果の文字列を出力する。

【0104】図12は本実施形態に係る自動ダイヤリング機能付き電話機の構成を示しており、主制御部50は、CPUで構成され、ROM51内に格納される所定の動作プログラムを実行することによりこの電話機の全体の動作を制御する。RAM52は、主制御部50で動作プログラムを実行するときに必要なデータを格納するとともに、主制御部50のための一時的なワーキングメモリとして用いられる。表示部53は、例えば液晶表示装置(LCD)等の表示装置であり、当該電話機の動作状態を表示したり、送信先の名称や電話番号を表示する。また、操作部54は、当該電話機を操作するために必要な文字キー、ダイヤル用テンキー、短縮ダイヤルキーや各種のファンクションキー等を備える。さらに、ネットワークコントロールユニット(NCU)55は、アナログの公衆電話回線Lの直流ループなどの閉結及び開

放の動作を行い、かつ自動ダイヤル機能を有するMTDFダイヤラーを含むハードウェア回路であり、必要に応じて送受話器59に接続し、もしくは音声合成出力部56からの出力を公衆電話回線L1に接続する。またさらに、音声合成出力部56は、例えば、パルス発生器と雑音発生器と利得可変型増幅器とフィルタとを備え、公知の音声合成方法を用いて、主制御部50からの制御により、音声合成すべき文字列のテキストデータを所定のパラメータ時系列に変換した後、そのピッチに基づいてパルス発生器を制御し、有声/無声の切り換えに基づいてパルス発生器と雑音発生器とを選択的に切り換えて使用し、振幅値に基づいて利得可変型増幅器を制御し、フィルタ係数値に基づいて上記フィルタを制御することにより、上記文字列の音声を生合成してスピーカ57を介して出力し、もしくは、当該音声合成の音声信号をNCU55及び公衆電話回線L1を介して通信の相手方に対して送信する。以上の回路51乃至56及び電話番号検索部60とは、バス58を介して主制御部50に接続される。

【0105】電話番号テーブルメモリ61は、人名とそれに対応する電話番号をテーブルの形式で予め記憶する。そして、電話番号検索部60は、音声認識装置100からの音声認識結果の「発信」という単語に続く文字列の人名の単語に基づいて、当該人名に対応する電話番号の情報を電話番号テーブルメモリ61から読み出して、当該電話番号の情報をバス58を介して主制御部50に出力する。これに回答して、主制御部50は、電話番号の情報をNCU55内のMTDFダイヤラーに出力し、このとき、NCU55は発信のためにオフフックした後、MTDFダイヤラーは入力される電話番号の情報に対応するダイヤル信号を発生して公衆電話回線L1に対して送出する。これにより、ユーザがマイクロホン1を介して発声した人名に対応する電話番号の電話機に対して発信できる。

【0106】この第2の実施形態によれば、音声認識及び自動ダイヤリング機能を備えた電話機において、単語辞書において未登録の未登録語に関する音声認識の精度を従来例に比較して高くすることができる音声認識装置を用いて情報検索を実行することができ、これにより、限られたメモリで多数の人名などの固有名詞が音声認識可能となるため、データベースを備えた電話機において、従来技術に比較して高い精度で情報検索が可能となる。また、高い音声認識率で自動ダイヤリングできる。

【0107】以上の実施形態においては、有限状態オートマトンモデル生成部23a及び有限状態オートマトンメモリ43aを備えているが、本発明はこれに限らず、それぞれ図1の単語クラスN-gramモデル生成部23、図1の単語クラスN-gramモデルメモリ43を備えてもよい。

【0108】＜第3の実施形態＞図13は、本発明に係

る第3の実施形態である構内交換機(PBX)の構成を示すブロック図である。この実施形態は、図11の音声認識装置100及び図12の電話番号検索部60、電話番号テーブルメモリ61を、構内交換機の内線転送又は外線転送に適用したことを特徴としている。

【0109】図13において、主制御部150は、CPUで構成され、ROM151内に格納される所定の動作プログラムを実行することによりこの構内交換機の全体の動作を制御する。RAM152は、主制御部150で動作プログラムを実行するときに必要なデータを格納するとともに、主制御部150のための一時的なワーキングメモリとして用いられる。表示部153は、例えば液晶表示装置(LCD)等の表示装置であり、当該構内交換機の動作状態を表示したり、送信先の名称や電話番号を表示する。また、操作部154は、当該構内交換機を操作するために必要な文字キー、ダイヤル用テンキーや各種のファンクションキー等を備える。さらに、ネットワークコントロールユニット(NCU)155は、外線である複数の公衆電話回線L1乃至LNと、内線電話機T1乃至TMに接続された内線とを交互に接続する電話交換スイッチ回路を備えるとともに、アナログの各公衆電話回線L1乃至LNの直流ループなどの閉結及び開放の動作を行い、かつ自動ダイヤル機能を有するMTDFダイヤラーを含むハードウェア回路であり、必要に応じて音声合成出力部156からの出力を公衆電話回線L1乃至LNに接続する。またさらに、音声合成出力部156は、図12の音声合成出力部56と同様に構成され、主制御部150からの制御により、音声合成すべき文字列のテキストデータを音声合成して、上記文字列の音声を生合成してその音声信号をNCU155及び公衆電話回線L1乃至LNを介して通信相手方に対して送信する。以上の回路151乃至156及び電話番号検索部60とは、バス158を介して主制御部150に接続される。

【0110】電話番号テーブルメモリ61bは、人名とそれに対応する内線電話番号及び外線電話番号をテーブルの形式で予め記憶する。そして、主制御部150は、NCU155において公衆電話回線L1乃至LNにうちの1つからの着信に対して自動応答し、「こちらは、ABC会社です。内線のどちらにお繋ぎしましょうか?」という音声合成信号を音声合成出力部156で発生させて相手方に出力する。これに対して、相手方から発声される内線転送すべき人名の音声信号をNCU155から音声認識装置100の特徴抽出器2に出力する。このとき、音声認識装置100は音声認識処理を実行し、音声認識結果の文字列を電話番号検索部60に出力する。これに回答して、電話番号検索部60は、音声認識装置100からの音声認識結果の文字列の人名の単語に基づいて、当該人名に対応する内線電話番号の情報を電話番号テーブルメモリ61bから読み出して、当該内線電話番

号の情報をバス158を介して主制御部150に出力する。これにตอบสนองして、主制御部150は、内線番号の情報に基づいてNCU155を制御して、当該着信してきた公衆電話回線に対応する内線番号の内線電話機に接続することにより、内線転送が完了する。

【0111】以上の実施形態においては、内線転送の例について説明しているが、内線電話機T1乃至TMから公衆電話回線への外線発信でも同様に、音声認識装置100、電話番号検索部60及び電話番号テーブルメモリ61bを用いて、ユーザが外線発信したい「人名」を発

声するだけで外線発信を実行できるように構成できる。
【0112】この第3の実施形態によれば、音声認識及び自動転送機能を備えた構内交換機において、単語辞書において未登録の未登録語に関する音声認識の精度を従来例に比較して高くすることができる音声認識装置を用いて情報検索を実行することができ、これにより、限られたメモリで多数の人名などの固有名詞が音声認識可能となるため、データベースを備えた構内交換機において、従来技術に比較して高い精度で情報検索が可能となる。また、高い音声認識率で自動転送できる。

【0113】<第4の実施形態>図14は、本発明に係る第4の実施形態であるカーナビゲーションシステムの構成を示すブロック図である。この実施形態は、図11の音声認識装置100をカーナビゲーションシステムに適用したことを特徴としている。

【0114】図14において、主制御部250は、CPUで構成され、ROM251内に格納される所定の基本プログラム及びCD-ROMドライブ装置259内のCD-ROMからフラッシュメモリ258にロードされたアプリケーションプログラムを実行することによりこのカーナビゲーションシステムの全体の動作を制御する。RAM252は、主制御部250で基本プログラム又はアプリケーションプログラムを実行するときに必要なデータを格納するとともに、主制御部250のための一時的なワーキングメモリとして用いられる。表示部253は、例えば液晶表示装置(LCD)等の表示装置であり、当該カーナビゲーションの動作状態を表示したり、指示された地名付近の地図やナビゲーション情報を表示する。また、操作部254は、当該ナビゲーションシステムを操作するために必要な文字キー、ダイヤル用テンキーや各種のファンクションキー等を備える。さらに、音声合成出力部256は、図12の音声合成出力部56と同様に構成され、主制御部250からの制御により、音声合成すべき文字列のテキストデータを音声合成して、上記文字列の音声を音声合成してその音声信号をスピーカ257に出力する。CD-ROMドライブ装置259には、カーナビゲーションのためのアプリケーションプログラム及び地図情報などのカーナビゲーション情報を格納したCD-ROMが挿入され、これらの情報は当該CD-ROMからCD-ROMドライブ装置259

及びバス258を介してフラッシュメモリ258にロードされて利用される。以上の回路251乃至256、259及び地名検索部60aとは、バス258を介して主制御部250に接続される。

【0115】地名テーブルメモリ61aは、地名とそれに対応する位置情報(緯度や経度の情報)をテーブルの形式で予め記憶する。そして、ユーザが音声認識装置100のマイクロホン1に対して地名を発声したとき、音声認識装置100は音声認識処理を実行し、音声認識結果の文字列を地名検索部60aに出力する。これにตอบสนองして、地名検索部60aは、音声認識装置100からの音声認識結果の文字列の地名の単語に基づいて、当該地名に対応する位置情報を地名テーブルメモリ61aから読み出して、当該位置情報をバス258を介して主制御部250に出力する。これにตอบสนองして、主制御部250は、当該位置情報に基づいて、フラッシュメモリ258内の地図情報などのカーナビゲーション情報を検索して、検索された対応する情報を表示部253に表示するとともに、検索された音声情報を音声合成出力部256に出力することにより、スピーカ257から当該音声合成された音声を出力させる。

【0116】この第4の実施形態によれば、音声認識及びカーナビゲーション機能を有するカーナビゲーションシステムにおいて未登録の未登録語に関する音声認識の精度を従来例に比較して高くすることができる音声認識装置を用いて情報検索を実行することができ、これにより、限られたメモリで多数の地名などの固有名詞が音声認識可能となるため、データベースを備えたカーナビゲーションシステムにおいて従来技術に比較して高い精度で情報検索が可能となる。また、高い音声認識率で地名を音声認識でき、適切にカーナビゲーションの処理を実行できる。

【0117】以上の第2、第3及び第4の実施形態においては、電話機、構内交換機、カーナビゲーションシステムの例について説明しているが、本発明はこれに限らず、単語リストに対応する普通名詞の単語データとそれに対応する情報を含むデータベースメモリを記憶し、音声認識装置100から出力される音声認識結果の文字列をキーとして用いて、上記データベースの記憶装置に記憶されたデータベースから検索して、一致する単語データに対応する情報を上記データベースメモリから読み出して出力し、さらには、当該検索された情報に基づいて、所定の処理を実行することができる。

【0118】<第5の実施形態>図15は、本発明に係る第5の実施形態であるかな漢字変換装置の構成を示すブロック図であり、図1と同様のものについては同一の符号を付している。この実施形態に係るかな漢字変換装置は、キーボード71と、キーボードインターフェース72と、音素HMMメモリ11及び単語辞書メモリ12が接続された単語照合部4aと、バッファメモリ5と、

統計的言語モデルメモリ44が接続された単語仮説絞込部6とを備えて構成される。

【0119】ここで、単語辞書メモリ12は、図1の単語辞書生成部22により生成された単語辞書を記憶し、ここで、単語辞書は、学習用データメモリ30に記憶されたファイル（図1や図11に図示の、日本人姓ファイル30a、日本人名ファイル30b、地名ファイル30cに限らず、上述のように、外国人の姓と名、会社名、各種施設名、各種製品名などの単語を含んでもよい。）及びテキストデータメモリ31内のテキストデータの単語に対応する漢字表記の複数の単語データを含む。また、統計的言語モデルメモリ44は、図1の言語モデル生成部24により生成された統計的言語モデルを記憶し、この統計的言語モデルは上記学習用データメモリ30に記憶されたファイル及びテキストデータメモリ31内のテキストデータの単語に基づいて生成される。

【0120】図15において、キーボード71は、かな文字列を入力するための入力手段であり、キーボードインターフェース72はキーボード71を用いて入力されたかな文字列のデータを一旦格納した後、所定の信号変換などの処理を実行した後、単語照合部4aに出力する。単語照合部4aは、ワンパス・ビタビ復号化法を用いて、キーボードインターフェース72を介して入力されるかな文字列のデータに基づいて、音素HMMメモリ11内の音素HMMと、単語辞書メモリ12内の単語辞書とを用いて単語仮説を検出し尤度を計算して出力する。ここで、具体的には、単語照合部4aは、単語辞書を参照して、入力されたかな文字列と、上記単語辞書内の単語との間の単語照合及び尤度計算を行い、一致したときに漢字表記の単語に変換して単語仮説の文字列として尤度とともに出力する一方、一致しないときにかな文字のまま単語仮説の文字列として尤度とともに出力する。単語照合部4aからの出力データはバッファメモリ5を介して単語仮説絞込部6に入力される。単語仮説絞込部6は、単語照合部4aからバッファメモリ5を介して出力される単語仮説に基づいて、統計的言語モデルメモリ44内の統計的言語モデルを参照して、終了時刻が等しく開始時刻が異なる同一の単語の単語仮説に対し、当該単語の先頭音素環境毎に、発声開始時刻から当該単語の終了時刻に至る計算された総尤度のうちの最も高い尤度を有する1つの単語仮説で代表させるように単語仮説の絞り込みを行った後、絞り込み後のすべての単語仮説の単語列のうち、最大の総尤度を有し漢字表記を含む仮説の単語列を認識結果として出力する。

【0121】この第5の実施形態によれば、かな漢字変換装置によれば、上記統計的言語モデルを利用して、かな漢字変換率を従来技術に比較して向上できるかな漢字変換装置を提供することができる。従って、例えば未登録の固有名詞も変換可能とすることができる。

【0122】

【発明の効果】以上詳述したように本発明によれば、クラスに依存して構築された未登録語モデルを含む統計的言語モデルを生成したので、次の特有の効果を得ることができる。

(1) モデル化対象を限定することで、読みの統計的特徴をより明確化することができ、クラス固有のパラメータ制約を導入できるため、未登録語モデルを高精度化することができる。

(2) 検出区間の言語処理が可能である。未登録語は、読みに加えクラスも同時に同定される。読みとクラスは、固有名詞の言語処理において必要十分な情報となるケースが多いものと考えられる。

(3) 上記生成された統計的言語モデルを用いて音声認識することにより、従来技術に比較して高い認識率で音声認識することができる。

【0123】また、本発明に係る情報検索処理装置によれば、電話機における音声認識及び自動ダイヤリング機能や、カーナビゲーションなどの小規模の情報検索処理装置において、単語辞書において未登録の未登録語に関する音声認識の精度が従来例に比較して高い音声認識装置を用いて情報検索を実行することができる。従って、限られたメモリで多数の人名や地名などの固有名詞が音声認識可能となるため、データベースを備えた小型携帯装置などの情報検索処理装置において、従来技術に比較して高い精度で情報検索が可能となる。

【0124】さらに、本発明に係るかな漢字変換装置によれば、上記統計的言語モデルを利用して、かな漢字変換率を従来技術に比較して向上できるかな漢字変換装置を提供することができる。従って、例えば未登録の固有名詞も変換可能とすることができる。

【図面の簡単な説明】

【図1】 本発明に係る第1の実施形態である連続音声認識システムのブロック図である。

【図2】 図1の連続音声認識システムにおける単語仮説絞込部6の処理を示すタイミングチャートである。

【図3】 図1の未登録語モデル生成部20によって実行される未登録語モデル生成処理を示すフローチャートである。

【図4】 図3のサブルーチンであるサブワード2-gramの単位決定処理（ステップS4）を示すフローチャートである。

【図5】 図1の言語モデル生成部24によって実行される言語モデル生成処理を示すフローチャートである。

【図6】 本発明者の分析による、日本人の姓及び名並びに旅行会話における単語の長さの分布を示すグラフであって、モーラ長に対する単語数の割合を示すグラフである。

【図7】 第1の実施形態に係るクラス依存未登録語モデルに基づく統計的言語モデルの一例を示す状態遷移図である。

【図8】 第1の実施形態に係る統計的言語モデルの一例を示す状態遷移図である。

【図9】 図1の未登録語モデル生成部20によって実行される未登録語モデル生成処理における、モーラ連鎖の単位化による平均尤度の向上を示すグラフであって、モーラ連鎖の種類の数に対する平均尤度を示すグラフである。

【図10】 本発明者による第1の実施形態の連続音声認識システムに係る実験の実験結果であって、日本人の姓及び名の再現率におけるモーラ連鎖の単位化効果を示すグラフであり、モーラ連鎖の種類の数に対する単語再現率を示すグラフである。

【図11】 本発明に係る第2の実施形態である連続音声認識システムの構成を示すブロック図である。

【図12】 図11の連続音声認識システムを用いた、自動ダイヤリング機能付き電話機の構成を示すブロック図である。

【図13】 本発明に係る第3の実施形態である構内交換機(PBX)の構成を示すブロック図である。

【図14】 本発明に係る第4の実施形態であるカーナビゲーションシステムの構成を示すブロック図である。

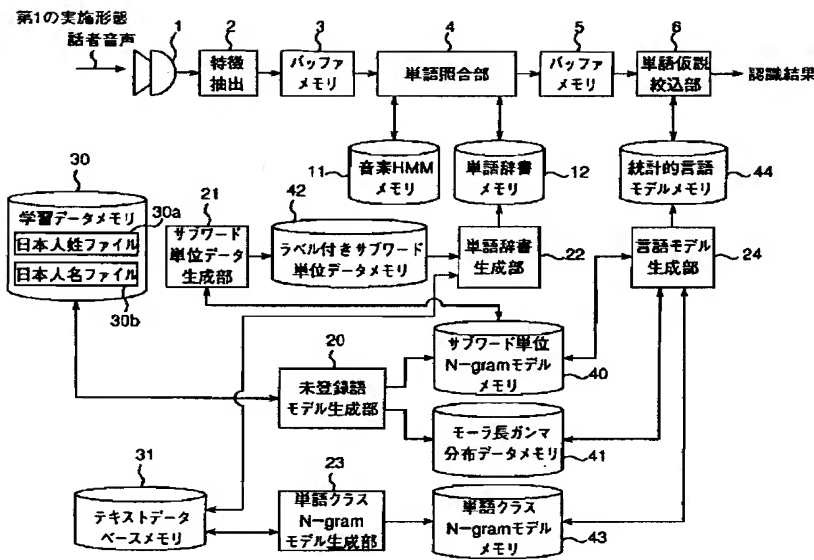
【図15】 本発明に係る第5の実施形態であるかな漢字変換装置の構成を示すブロック図である。

【符号の説明】

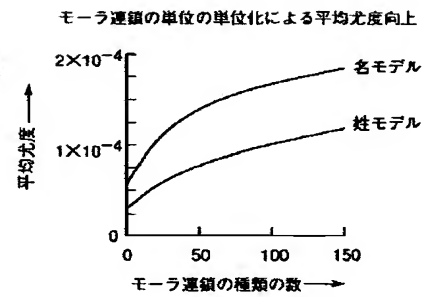
1…マイクロホン、
2…特徴抽出部、
3、5…バッファメモリ、
4、4a…単語照合部、
6…単語仮説絞込部、
11…音素HMMメモリ、
12…単語辞書メモリ、
20…未登録モデル生成部、
21…サブワード単位データ生成部、
22…単語辞書生成部、
23…単語クラスN-gramモデル生成部、
23a…有限状態オートマトンモデル生成部、
24…言語モデル生成部、
30…学習データメモリ、
30a…日本人姓ファイル、
30b…日本人名ファイル、
30c…地名ファイル、
31…テキストデータベースメモリ、
40…サブワード単位N-gramモデルメモリ、

41…モーラ長ガンマ分布データメモリ、
42…ラベル付きサブワード単位データメモリ、
43…単語クラスN-gramモデルメモリ、
43a…有限状態オートマトンモデルメモリ、
44…統計的言語モデルメモリ、
50…主制御部、
51…ROM、
52…RAM、
53…表示部、
54…操作部、
55…ネットワークコントロールユニット(NCU)、
56…音声合成出力部、
57…スピーカ、
58…バス、
59…送受話器、
60…電話番号検索部、
60a…地名検索部、
61、61b…電話番号テーブルメモリ、
61a…地名テーブルメモリ、
71…キーボード、
72…キーボードインターフェース、
100…音声認識装置、
150…主制御部、
151…ROM、
152…RAM、
153…表示部、
154…操作部、
155…ネットワークコントロールユニット(NCU)、
156…音声合成出力部、
158…バス、
250…主制御部、
251…ROM、
252…RAM、
253…表示部、
254…操作部、
256…音声合成出力部、
257…スピーカ、
258…バス、
259…CD-ROMドライブ装置、
L、L1乃至LN…公衆電話回線、
T1乃至TM…内線電話機。

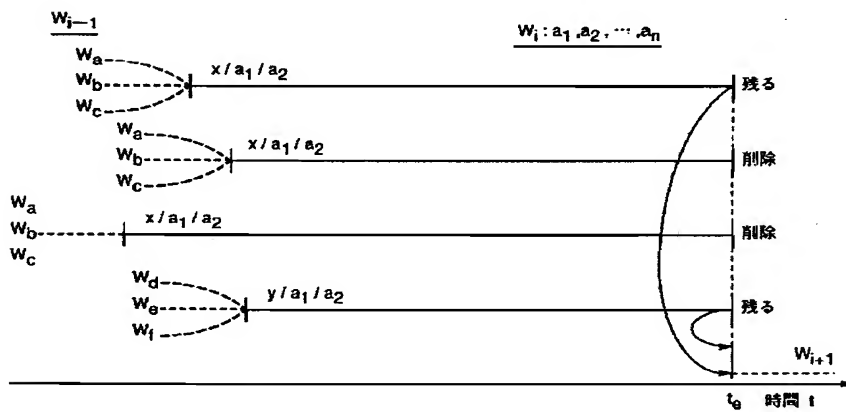
【図1】



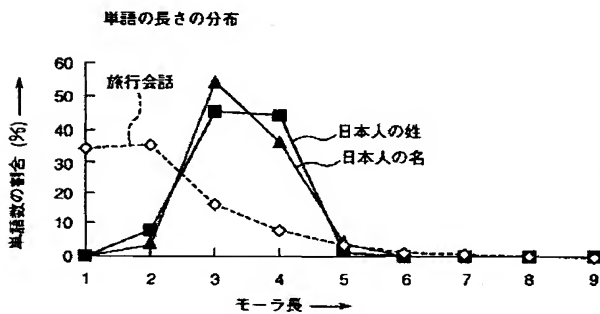
【図9】



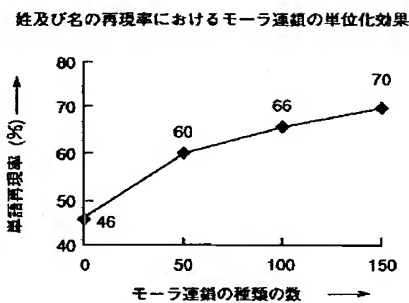
【図2】



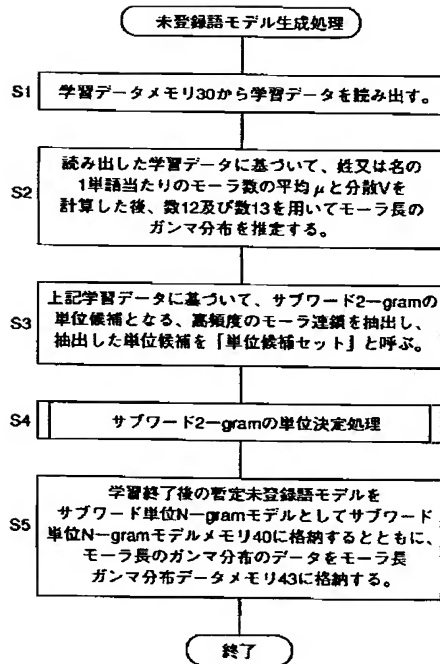
【図6】



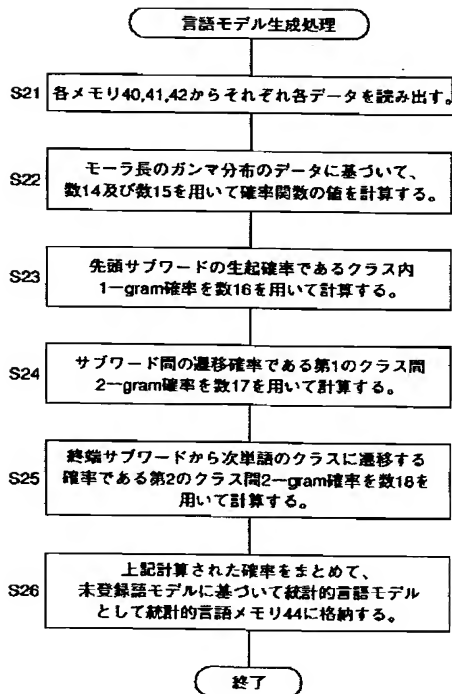
【図10】



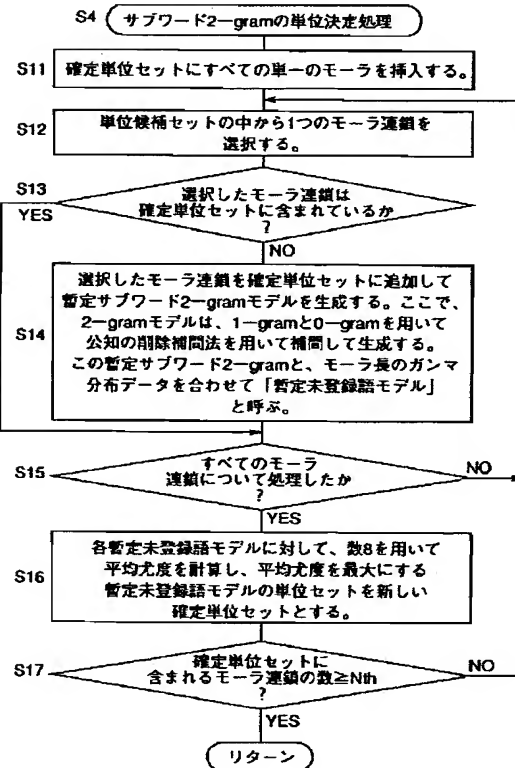
【図3】



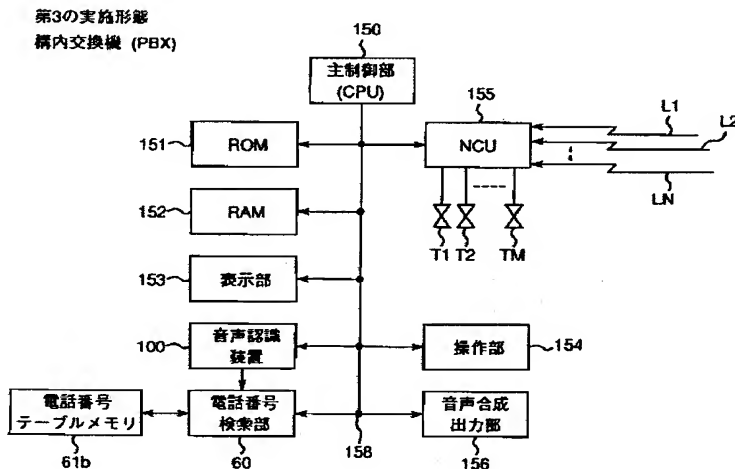
【図5】



【図4】

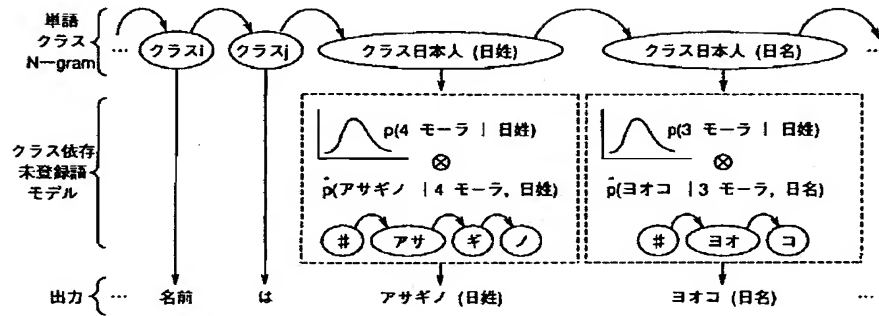


【図13】



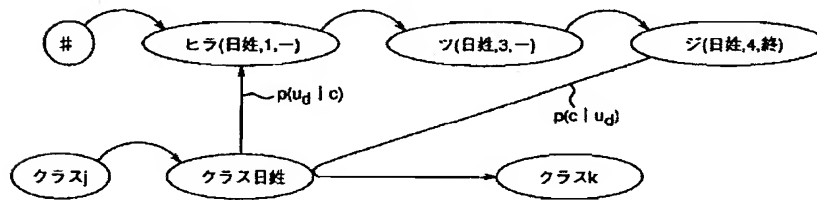
【図7】

クラス依存未登録語モデルに基づく統計的言語モデルの一例

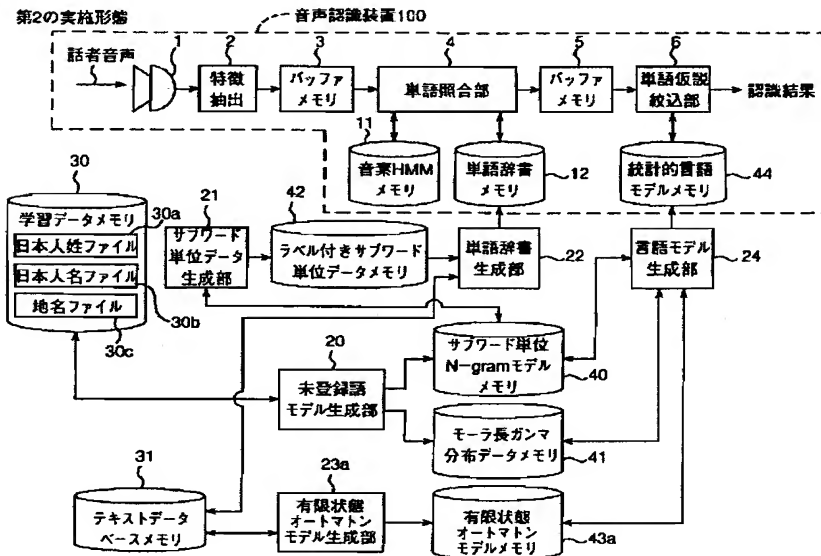


【図8】

統計的言語モデルの一例

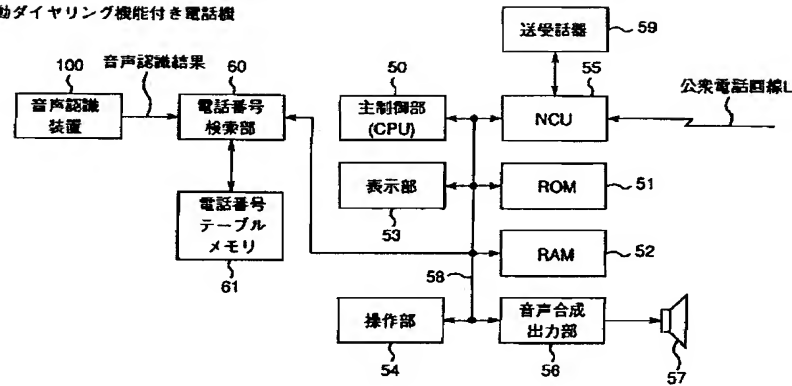


【図11】

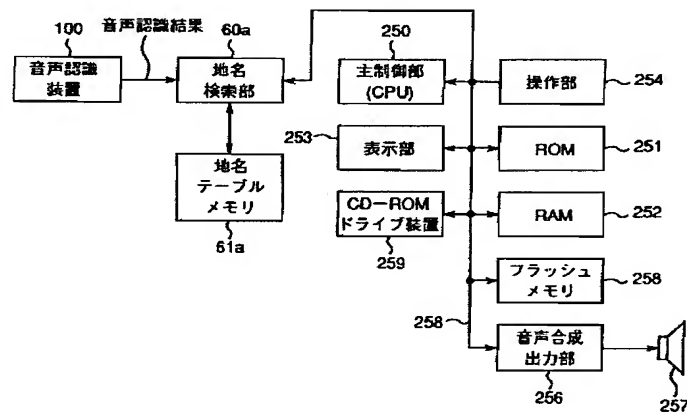


【図12】

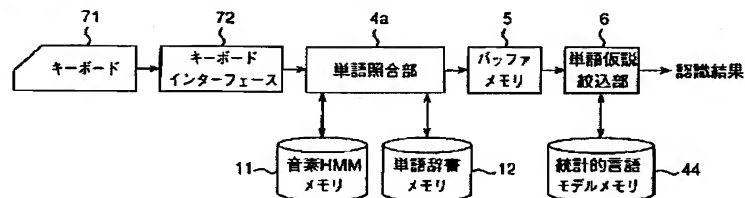
自動ダイヤリング機能付き電話機



【図14】

第4の実施形態
カーナビゲーションシステム

【図15】

第5の実施形態
かな漢字変換装置

フロントページの続き

(72)発明者 山本 博史

京都府相楽郡精華町大字乾谷小字三平谷5

番地 株式会社エイ・ティ・アール音声翻

訳通信研究所内

(72)発明者 匂坂 芳典

京都府相楽郡精華町大字乾谷小字三平谷5

番地 株式会社エイ・ティ・アール音声翻

訳通信研究所内